LEARNING, HYGIENE AND TRADITIONAL MEDICINE*

Daniel Bennett, Asjad Naqvi and Wolf-Peter Schmidt

Information provision is only an effective behaviour-change strategy if the information is credible. A novel programme augments conventional hygiene instruction by showing participants everyday microbes under a microscope. Through a randomised evaluation in Pakistan, we show that this programme leads to meaningful hygiene and health improvements, while instruction alone does not. Traditional medicine, which offers an alternative disease model, may undermine learning by strengthening prior beliefs about hygiene. We show that believers in traditional medicine have smaller impacts, suggesting that traditional and modern medical beliefs are substitutes and that traditional medicine may exacerbate the infectious disease burden in this context.

Information provision is a common behaviour-change strategy in many economic contexts. In the health field, policymakers rely on information provision to prevent HIV, discourage smoking, improve nutrition and encourage chronic disease compliance. Interventions range from cigarette warning labels to community health worker programmes (WHO 2008; Kamyab *et al.*, 2014). However, impact evaluations of information provision in public health show mixed effects on risky sexual behaviour (De Walque, 2007; Dupas, 2011), nutrition (Avitabile, 2012; Luo *et al.*, 2012; Wong *et al.*, 2014), malaria prevention (Rhee *et al.*, 2005), and sanitation (Cairncross *et al.*, 2005; Madajewicz *et al.*, 2007). While Dupas (2011) shows that girls in Kenya select younger (and safer) partners after learning age-specific rates of HIV prevalence, Luo *et al.* (2012) and Wong *et al.* (2014) show that nutrition education for parents in China does not reduce child anaemia. Guiteras *et al.* (2014) find that neither messages appealing to negative emotions or social pressure improve hand washing or increase the willingness to pay for water chlorination in Bangladesh.

Information must be convincing in order to change behaviour. Messages related to infectious disease prevention typically invoke the germ theory of disease, which states that infectious diseases are caused by microbes. This model presupposes the existence of microbes, which may not be self-evident to people with limited education. Many people throughout the world believe in systems of traditional medicine that feature non-pathogenic theories of disease. Unani medicine, the most common traditional medical system in Pakistan, does not conceive of invisible pathogens. Instead, an imbalance between humoural elements within the body leads to illness (Anwar *et al.*, 2012; Karmakar *et al.*, 2012). Diarrhoea is a symptom of overconsumption and 'excess heat', rather than an enteric infection. Standard hygiene messages to wash hands with soap, handle food safely, purify water and maintain latrines may not be credible to people who are unfamiliar with microbes or believe in traditional medicine.

Microbe Literacy (ML) is a novel hygiene (WASH) programme in rural Pakistan that attempts to make hygiene instruction more salient. Instructors use microscopes to demonstrate the presence of microbes in common substances like standing water, buffalo dung and spoiled food. Later, participants learn about appropriate hygiene practices, including hand washing, safe food handling and latrine usage, as well as the causes and consequences of infant diarrhoea. By showing that microbes exist, the microscope demonstration indirectly substantiates the germ theory of disease. Programmers hope that participants who have seen microbes directly will be more receptive to hygiene messages.

This study evaluates the impact of ML through a cluster-randomised trial. We offered the programme to female adult literacy class (ALC) participants in southern Punjab Province, Pakistan. One treatment arm received ML, another arm received only hygiene instruction, and a third arm received no programming. This design allows us to assess the absolute impact of ML and isolate the contribution of the microscope demonstration. We consider respondent hygiene and a primary outcome and household sanitation and respondent and child health as secondary outcomes. Throughout the analysis, we use the Romano and Wolf (2005) stepdown procedure to adjust p-values for multiple hypothesis testing across outcomes. ML improves our primary hygiene proxy by 0.25 standard deviations in the three-month midline survey. The midline health impact is mixed, with a strong effect on self-reported morbidity for respondents and a weaker effect for children. Impacts persist and strengthen by the 16-month endline survey, in which we also find effects on household sanitation, hygiene of other household members, and child anthropometrics. In contrast, hygiene instruction alone has little or no effect on any outcomes.

In a heterogeneity analysis, the article considers the relationship between hygiene education and traditional medical beliefs. We define a traditional belief index (TBI) by summing four indicators of the belief in Unani medicine. We interact baseline values of the index with treatment and find that traditional beliefs substantially weaken the impact of ML. The impact on hygiene is three times larger for participants with weak traditional beliefs than for those with strong beliefs. The entire anthropometric effect occurs among children of participants with weak beliefs. These results are robust across four alternative TBI constructions. A reasonable concern is that estimates could reflect other omitted determinants of learning, such as cognitive ability or socio-economic status. We control for the interaction between treatment and a battery of baseline variables, including demographic and economic characteristics, hygiene, sanitation, and health measures, and ALC test scores. The robustness to these controls suggests that the result is not spurious. We also show that being uninformed about hygiene generally does not lead to a similar heterogeneous response. Finally, we show the impact of the intervention on traditional medical beliefs. While ML reduces the TBI in the midline survey, this effect disappears by the endline survey.

We sketch a simple learning model below (developed formally in the Appendix) to motivate and interpret these results. In a Bayesian framework, the mean and precision of the prior and the signal determine the impact of information. By showing the existence of microbes, the microscope demonstration may increase the precision (i.e. the credibility) of the signal. However, as we discuss below, it could also increase the signal mean or increase the signal precision through other channels. Traditional

medical beliefs may attenuate the impact of information by increasing the prior precision. Alternatively, traditional beliefs could strengthen the impact of information by reducing the prior mean. The negative interaction between ML and traditional medical beliefs suggests that traditional beliefs influence the prior precision.

This study makes two primary contributions. Using a rigorous, randomised design, we show that an intervention designed to educate participants about microbes has a strong, lasting impact while instruction alone does not. Notwithstanding several caveats, this finding suggests that the provision of corroborating evidence (which is not the default approach in this area) may strengthen the impact of health recommendations. It may also help to explain heterogeneity across studies in the impact of hygiene promotion (Curtis and Cairncross, 2003; Fewtrell *et al.*, 2005; Aiello *et al.*, 2008; Waddington and Snilstveit, 2009). Our findings also contribute to a broader literature on learning and technology adoption (Conley and Udry, 2010; Argent *et al.*, 2014).

Secondly, this article is one of the first well-identified analyses of the role of traditional medical beliefs. We uncover substantial heterogeneity in the impact of information according to beliefs in traditional medicine. Despite the ubiquity of traditional medicine (WHO 2003), very little social science research considers its implications (Leonard, 2003; Leonard and Graff Ziven, 2005; Wang *et al.*, 2010; Kooreman and Baars, 2012; Sato, 2012*a,b*). Traditional medical beliefs may help to explain the low demand for ostensibly valuable health products like deworming treatment, antimalarial bed nets, and low-emission cook stoves (Kremer and Miguel, 2007; Cohen and Dupas, 2010; Mobarak *et al.*, 2012). Although it is difficult to generalise, results suggest that traditional beliefs exacerbate the burden of infectious disease in this setting by discouraging healthy behaviour.

1. Background

1.1. *Theoretical Motivation*

This subsection motivates the experiment theoretically and discusses the interaction between hygiene information and traditional medical beliefs. We draw upon a formal model of Bayesian learning and hygiene behaviour, which appears in Section A1 of the online Appendix. Suppose that health is a function of hygiene and another (composite) health input. Although the marginal product of hygiene is not directly observable, people use available information to formulate beliefs about this parameter. Before the intervention, people have priors about the health impact of hygiene. The mean of the prior represents the perceived impact of hygiene and the precision (inverse variance) represents the certainty of this perception.

Hygiene education provides an informational signal about the marginal product of hygiene. The mean and precision of the signal represent the strength and credibility of the message. We assume that the signal mean exceeds the prior mean, so that people receive the message that hygiene is more valuable than they thought. We assume for simplicity that people update their beliefs according to Bayes' rule. Other learning mechanisms are also possible, and our analysis does not attempt to distinguish among alternative learning models. Under a standard utility function and a linear budget constraint, an increase in the perceived marginal product of hygiene also increases hygiene and health.

The experiment compares ML to an intervention without the microscope demonstration ('Instruction Only') and a control group. The main rationale for the microscope demonstration is to make subsequent hygiene messages more convincing, which equates to an increase in the signal precision. As we discuss in subsection 3.4, the intervention could improve the signal precision through channels other than awareness of microbes. It could also increase the signal mean if the microscope demonstration indirectly provides hygiene information. Regardless of the channel, the microscope demonstration should strengthen the impact of hygiene education.

Traditional medical beliefs may moderate the impact of the intervention by influencing priors about the marginal product of hygiene. Unani medicine offers an alternative disease model in which the marginal product of hygiene is low. Adherence to this system may have two distinct effects on participants' priors. Traditional medical beliefs may increase the prior precision by offering a rationale for the belief that hygiene is ineffective. This channel leads believers to place less weight on the informational signal than non-believers. Traditional medical beliefs may also decrease the prior mean, so that believers learn more because they have more potential to learn. The sign of the interaction between treatment and traditional medical beliefs is theoretically ambiguous because these channels work in opposite directions. However, a negative interaction only arises in the model if traditional beliefs influence the prior precision.

1.2. *Information, Hygiene and Health*

Infectious diseases such as diarrhoea and respiratory infections are a primary cause of child mortality and morbidity in developing countries (Kosek *et al.*, 2003; WHO 2013*a*,*b*). Diarrhoea may lead to death through dehydration, which is the rationale for the standard modern treatment of oral rehydration therapy. Mortality occurs infrequently relative to the prevalence of diarrhoea, which is around 32% within the past two weeks for children in our sample. Chronic diarrhoea also interferes with nutrition and cognitive development (Guerrant *et al.*, 1999; Niehaus *et al.*, 2002). Environmental enteropathy is a subclinical disorder in which frequent intestinal infections lead to chronic malabsorption of nutrients, which in turn interferes with physical growth (Langford *et al.*, 2011). A meta-analysis by Checkley *et al.* (2008) shows that the odds of stunting increase by 1.13 for every five diarrhoea episodes. Morbidity among adults also influences child health and development. Illnesses for pregnant women may directly hamper foetal development (Almond, 2006; Lin and Liu, 2014). More generally, parental health shocks may affect human capital investment though several channels (Gertler and Gruber, 2002; Bratti and Mendola, 2014).

Efforts to address diarrhoea in developing countries usually involve either infrastructure or communication. Infrastructure projects include latrine construction and water source protection (Kremer *et al.*, 2011). Since these projects are relatively expensive, policymakers have sought effective behaviour-change interventions such as education and community-led total sanitation (Pattanayak *et al.*, 2009). A rich public health literature evaluates the impact of 'hygiene promotion' on diarrhoea and child anthropometrics (Curtis and Cairncross, 2003; Fewtrell *et al.*, 2005; Aiello *et al.*, 2008; Ejemot-Nwadiaro *et al.*, 2008; Waddington and Snilstveit, 2009). These interventions typically bundle information with subsidies such as free soap, making it difficult to

isolate the role of learning. Although they are often time-limited (Vindigni *et al.*, 2011), many of these studies find significant effects.

Information provision is common in many economic contexts, including health promotion and technology adoption.[1] In health, there is a distinction between programmes that offer personalised and general information. Tailored messages, such as HIV test results and household water contamination reports, appear to change behaviour (Madajewicz *et al.*, 2007; Jalan and Somanathan, 2008; Thorton, 2008). However, the evidence regarding general prevention messages is mixed. For instance, Bowen *et al.* (2012) find an impact of a nine-month hygiene promotion campaign in Pakistan, while Luo *et al.* (2012) and Wong *et al.* (2014) show that nutrition information does not change nutrition and anaemia for children in China. Guiteras *et al.* (2014) show that message framing does not influence hand washing or the willingness to pay for water chlorination in Bangladesh.

Increasing the salience of the germ theory is a novel approach. To our knowledge, a community-based intervention by Ahmed and Zeitlan (1993) in rural Bangladesh is the closest analogue to our study.[2] In that study, which uses a non-randomised design, participants met at least weekly over six months in small groups and were encouraged to improve sanitation and hygiene. In addition to other activities, facilitators demonstrated the existence of microbes through a chemical reaction that caused the microbes suspended in water to precipitate. The study finds significant improvements in hygiene and health, as well as an impact of 0.28 points on weight-for-age after 12 months, which is very similar to our estimate.

1.3. *Traditional Medicine*

Traditional medicine is ubiquitous, comprising up to 50% of health care utilisation in China and up to 80% of utilisation in sub-Saharan Africa (WHO 2003). Forty per cent of US adults use at least one form of complementary or alternative medicine (CAM) (Barnes *et al.*, 2008). *Unani Tibb* ('Greek Medicine' in Urdu) is a common form of traditional medicine in South Asia. In this system, disease causes and treatments are based on the four humours of blood, mucus, yellow bile and black bile, which combine with the four qualities of heat, cold, moisture, and dryness (Sheehan and Hussain, 2002). Since humoural imbalances are believed to cause illness, Unani treatments adjust exposure to humoural elements in order to restore balance (Mull and Mull, 1988). Objects, foods, and actions have 'hot' and 'cold' designations that do not correspond to physical temperature. Although designations vary, lamb, eggs, lemons, olives, ginger, cinnamon and honey are considered hot while beef, okra, banana, melon, and vinegar are considered cold. Diarrhoea is a symptom of excess heat and overconsumption in the Unani system, although some people also perceive it as a cold disease (Nielsen *et al.*, 2003).

---

[1] A literature in economics examines the role of learning in agricultural technology diffusion (Foster and Rosenzweig, 1995; Conley and Udry, 2010; Argent *et al.*, 2014). Genius *et al.* (2014) and Krishnan and Patnam (2014) examine the impact of agricultural extension services on technology adoption. The credibility of extension agents has parallels to the credibility of hygiene education facilitators in this study.

[2] Ahmed *et al.* (1993, 1994) also discuss this evaluation.

Traditional and modern medicine may be complements or substitutes. Someone can believe that both humoural imbalances and microbes cause disease, or that imbalances increase the susceptibility to infection. Many people utilise both traditional and modern providers for the same illness (Hunte and Sultana, 1992), which may cause substitution through the budget constraint. The interaction between traditional and modern medicine is a critical public health question since beliefs in traditional medicine could lead ineffective approaches to crowd out effective approaches. Traditional medicine is diverse, and this concern clearly varies by setting, disease, and belief system. Perhaps surprisingly, the World Health Organisation promotes traditional medicine because it is 'accessible and affordable' despite the lack of evidence of general effectiveness (WHO 2013a,b).

In the case of diarrhoea, the Unani and modern approaches directly conflict; whereas oral rehydration therapy involves giving liquids, the standard Unani treatment for diarrhoea is to consume fewer liquids and foods, especially if they are hot (Mull and Mull, 1988; Nielsen *et al.*, 2003). In particular, a nursing mother who has been working outdoors should withhold breast milk, which is considered hot, from an infant with diarrhoea.[3] This tension means that traditional medical beliefs may be an important moderator of the impact of hygiene instruction.

2. Study Design

2.1. *Description of the Intervention*

We conducted this study in rural villages in southern Punjab Province, Pakistan. Wheat and cotton cultivation are the main economic activities and Sunni Islam is the dominant religion in this area. Communities are culturally conservative and practice Purdah, which severely limits female autonomy and mobility. However, observance of Islamic customs such as daily prayer and Ramadan varies with the level of poverty. People generally value cleanliness but focus on dirt that they can perceive through sight or smell (Nielsen *et al.*, 2003). Women typically marry by age 20.

The National Commission for Human Development (NCHD) is a non-profit organisation that collaborates with the national government to provide health care and education in poor communities. The NCHD regularly conducts ALCs throughout Pakistan for women without formal schooling. Officials obtain the approval of community leaders before offering instruction in a community. Classes, which are free, meet for 90 minutes, six days per week for six months in the home of a local volunteer. The curriculum covers basic literacy and numeracy: students learn to read and write the alphabet, form simple sentences, and perform basic arithmetic. Students sit for three literacy tests and one mathematics test during the ALC.

Microbe Literacy is a hygiene information programme developed by the Microbe Literacy Initiative (MLI, formerly the South Asia Fund for Health and Education), an

---

[3] Traditional and modern approaches to HIV/AIDS conflict in a similar way. Thirty per cent of respondents in the 2008 Ghana DHS believe that witchcraft causes AIDS (Tenkorang *et al.*, 2011), which may lead people to deemphasise avoiding risky sex (Yamba, 1997). Many people in sub-Saharan Africa also subscribe to the 'virgin cleansing myth' that sex with a virgin is a cure for AIDS. This belief directly encourages risky behaviour.

international NGO. The programme includes a microscope demonstration and an infection prevention workshop, which each take 90 minutes and occur several days apart. Section A2 in the online Appendix provides the complete curriculum for both ML components. The MLI and the NCHD have collaborated to provide ML as a pilot programme in several settings in Pakistan. For this intervention, the NCHD recruited existing employees to be ML instructors. Instructors were chosen from nearby communities based on education and job performance. Individual instructors served in all treatment arms that received instruction (the ML and IO arms, described below), so that there was no distinction by treatment arm in instructor characteristics. For ML ALCs (which received both the microscope demonstration and the infection prevention workshop), the same instructors generally provided both intervention components.

Participants in the microscope demonstration use a microscope to view the microbes in their environment. To begin, facilitators distribute magnifying glasses and explain the concept of magnification. Participants learn that both a magnifying glass and a microscope make small objects appear bigger. Next, facilitators and participants create microscope slides of everyday substances from nearby, such as standing water, buffalo dung, and spoiled food. Participants take turns looking through the microscope while their classmates observe on a monitor. The microscope demonstration does not include instruction related to hygiene, disease prevention, or diarrhoea treatment and does not directly promote the germ theory.

The infection prevention workshop provides strategies to avoid infectious disease. Instructors say that microbes are all around us, but that only some microbes cause disease. They stress that because microbes are invisible, our hands and drinking water could be contaminated even if they look clean. Participants learn to wash their hands with soap after defecating and before preparing food. They also learn to purify and protect drinking water sources, to maintain a clean cooking area, and to avoid contamination by flies. The lesson emphasises that children can die from diarrhoea, and recommends treating diarrhoea by giving uncontaminated liquids and foods, including breast milk for nursing infants.[4]

The experiment has three treatment arms: the ML arm received both the microscope demonstration and the infection prevention workshop, the Instruction Only (IO) arm received only the infection prevention workshop, and the Control (C) arm received no hygiene education through the programme. A comparison of ML and C shows the absolute impact of ML while a comparison of ML and IO shows the impact relative to a prescriptive hygiene lesson.

Although the study encompasses a large geographical area, some ALCs are nearby other ALCs. Indirect learning by control respondents could bias estimates downward. However, low female mobility and autonomy limit the scope for informational spillovers in this context. We reduced the potential for spillovers by combining ALCs into 110 randomisation groups that were at least one kilometre apart. After

---

[4] Field reports suggest that the programme left a strong impression on participants. One facilitator remarked: 'It was amazing for women to see the bacteria on the slides which had been sampled from their homes. Women were very astonished to know how much bacteria live around them. They expressed that they will be careful to avoid microbes for themselves and for their children'. In a pilot study in the Swat Valley of Pakistan, Ahmad *et al.* (2012) found that ML was associated with a 65% decline in diarrhoea and a 76% decline in respiratory illness. These results are difficult to interpret because the study lacks a control group.

randomising at this level, the median distance from control ALCs to the nearest ML or IO ALC is three kilometres. Section A7 in the online Appendix discusses the potential for control group contamination further. We stratified the sample across four districts and three diarrhoea prevalence categories, for a total of 12 strata, to encourage balance on spatial and health characteristics. This process led to 72 ALCs in the ML arm, 71 ALCs in the IO arm, and 68 ALCs in the C arm.

2.2. *Data and Measurement*

We collaborated with the NCHD to evaluate the impact of ML among ALC participants in 2013 and 2014. We selected 210 ALCs in southern Punjab Province and enrolled participants who were at least 15 years old. The baseline survey took place in May of 2013. The intervention occurred in June and July. We conducted a three-month midline survey in August and September of 2013 and a 16-month endline survey in October and November of 2014. Interviews occurred in respondents' homes to avoid the influence of ALC instructors and classmates. The endline survey included a limited survey of other female members of ALC participants' households, which we describe further below.

Our baseline sample includes 4,032 respondents. Thirty-two per cent of respondents have children age five and at baseline, for a total of 1,949 baseline children. Midline attrition was 5% for respondents and 7% for children. Endline attrition was 15% for respondents and 29% for children relative to baseline. A severe flood in Muzaffargarh District prevented surveyors from reaching 10 ALCs, leading to endline attrition for 6% of respondents and 9% of children. Since it was local to the affected ALCs, the flood did not otherwise affect the data collection.[5] We implement listwise deletion for 141 respondent observations and 55 child observations (across the midline and endline rounds) with missing values for any outcome variables in our analysis. We also exclude child observations with biologically implausible weight-for-age and height-for-age z-scores from anthropometric estimates. For outcome variables that were collected at both midline and endline, the sample includes 7,103 respondent observations and 3,945 child observations. Section A3 in the online Appendix discusses attrition and missing data in more detail.[6]

Adult literacy class participants are not representative of the community, and two sources of selection may affect the interpretation of the results below. Since the adult literacy programme targets women without formal schooling, participants may be relatively disadvantaged. Section A4 in the online Appendix compares the study sample to a representative sample of rural Punjab women. Only 11% of study participants have any formal schooling, compared to 42% of the representative sample. Study participants are 7.3 years younger than average and 10 percentage points less likely to be married.[7]

---

[5] Flood-affected respondents and other attriters have very similar baseline characteristics to respondents who remain in the sample, as we discuss further in the online Appendix.

[6] Child hygiene has a smaller sample size because it is recorded per respondent and many respondents do not have children. Anthropometric estimates are larger and more statistically significant if we include extreme observations.

[7] Section A4 in the online Appendix also estimates the interaction between treatment and the respondent characteristics that are available in both samples. We use these estimates and sample frequencies of these variables to predict the impact of ML on our main outcomes in both samples. This approach provides a suggestive indication of whether our results are sensitive to socio-economic status.

As seekers of human capital, ALC participants may have unusual levels of motivation, learning ability, or autonomy. One concern is that impacts could be weaker among people who have not self-selected in this way. This form of selection is likely to be mild in practice because the NCHD works hard to recruit all eligible women and offers as many classes as necessary. Two additional results limit this concern. Section A4 in the online Appendix shows that treatment effects do not depend on the level or change in ALC test scores, which measure academic ability and proxy for the utility of ALC participation. Second, Table 5 shows similar results for other female household members, who have not been selected based on ALC participation.

Participant hygiene is the primary outcome of this study. Hygiene is particularly difficult to measure. Hygiene encompasses several behaviours, including hand washing, latrine use, safe food handling and bathing. People do not provide reliable self-reports and direct observation of behaviour is not usually feasible (Ram, 2010). For instance, 96% of respondents report washing their hands after defecating although only 29% have hand washing stations. Structured observation, an alternative in which surveyors observe behaviour over longer intervals, is subject to Hawthorne effects. Hand rinse cultures may indicate faecal contamination of hands, but Ram *et al.* (2011) show that measurements are noisy and are only weakly correlated with behaviour.

We proxy for hygiene through direct observation of the personal appearance of respondents. Surveyors record the overall appearance on a 3-point Likert scale: 3 indicates good hygiene (no visible dirt on hands, feet, clothes, or fingernails), 2 indicates moderate hygiene (some visible dirt) and 1 indicates poor hygiene (extensive visible dirt). Several studies have validated this approach to hygiene measurement (Ruel and Arimond, 2002).[8] Figure 1 shows the baseline frequency distribution for hygiene: 41% of respondents have good hygiene, 55% have moderate hygiene, and 4% have poor hygiene. The endline survey includes separate measurements of the appearance of hands, fingernails, clothing, and feet. Section A5 provides several additional validity checks of this hygiene proxy.[9]

Household sanitation, child hygiene, and respondent and child health are secondary outcomes of this study. For respondents with children under age 5 (1,288 baseline respondents), surveyors assessed the overall appearance of children who were present during the interview.[10] They observed and recorded the extent of open defecation and garbage disposal on 4-point Likert scales, as well as the cleanliness of the cooking area on a 3-point scale.[11] They recorded whether the household had a handwashing station

[8] According to Ram (2010), 'While these indicators do not directly indicate handwashing behaviour, they are currently used as surrogate markers because they are reliable and efficient'. Luby *et al.* (2011) show that spot checks of hand cleanliness are correlated with child health. Halder *et al.* (2010) show that observed cleanliness is correlated with handwashing in structured observations.

[9] Section A5 in the online Appendix shows that personal appearance is strongly correlated with other hygiene and sanitation indicators at baseline. It also uses a bootstrap approach to show that the treatment effects on hygiene and health are strongly correlated. Finally, this section addresses a handful of other ways that personal appearance might not accurately reflect hygiene behaviour.

[10] The sample sizes for child hygiene and child health differ because the health observations are associated with individual children while hygiene observations relate to all children younger than 5 of the respondent. Table A2 in the online Appendix provides the sample sizes for all outcomes in the article.

[11] Absence of defecation has the following categories: (*i*) heavy defecation in the area; (*ii*) some defecation in the area; (*iii*) very little excreta visible; and (*iv*) no excreta visible. Absence of garbage has categories: (*i*) lots of uncollected garbage; (*ii*) some uncollected garbage; (*iii*) very little garbage; and (*iv*) no garbage visible. Cleanliness of the cooking area has categories: (*i*) filthy; (*ii*) not so clean; and (*iii*) very clean.
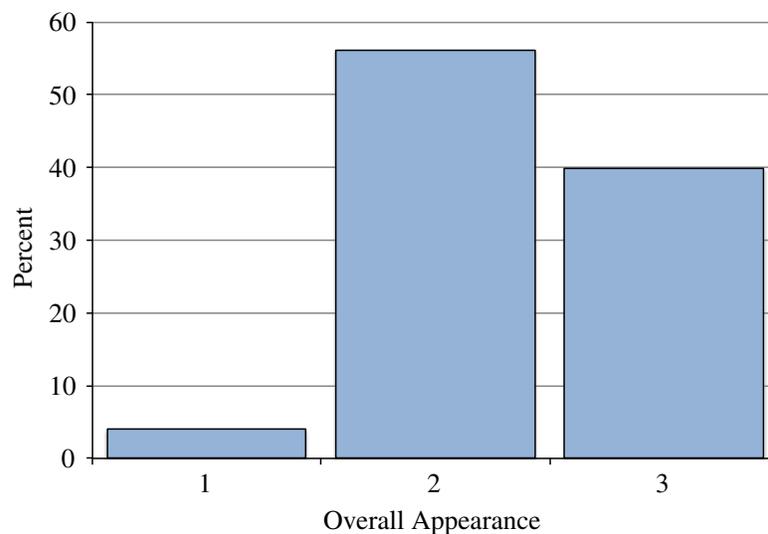
Fig. 1.   *Baseline Respondent Hygiene*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

with soap. Since these outcomes are public goods that are determined through intrahousehold bargaining, we may expect weak treatment effects if study participants lack bargaining power.

We measure health by eliciting the prevalence of diarrhoea, fever and cough for the respondent and her children younger than 5 within the past two weeks. While it is a key public health benchmark, self-reported morbidity is measured with considerable error (Schmidt *et al.*, 2010, 2011). People define these conditions in different ways and often forget instances of morbidity. Measurement error is likely to be worse for children than for adult respondents because adults who answer on behalf of children may be unaware of some instances of morbidity. We consolidate these outcomes into a health index for concision but report disaggregated estimates in Section A6 in the online Appendix. The health index is defined as the number of absent morbidities and ranges from 0 to 3. Estimates are similar if we weight morbidities according to a summary index (Anderson, 2008) or utilise the first principal component. Figures 2 and 3 show the baseline frequency distributions of the health index for respondents and children. Thirty-four per cent of respondents and 38% of children had at least one morbidity at baseline.

The endline survey includes weight and height measurements for children younger than 5. Gastrointestinal infections may affect child weight and height through environmental enteropathy, a subclinical condition in which frequent infections reduce nutrient uptake (McKay *et al.*, 2010). We convert weight and height into weight-for-age and height-for-age z-scores (WAZ and HAZ) using the WHO anthropometry database. Our analysis excludes children with WAZ or HAZ scores that are less than -6 or greater than 5, which the WHO considers biologically implausible. With a median WAZ of -1.71 and a median HAZ of -2.48, the sample includes many children who are severely malnourished. The anthropometric sample is smaller than the child health sample because we only observe weight and height at endline and because we limit the sample to children with biologically plausible z-scores.
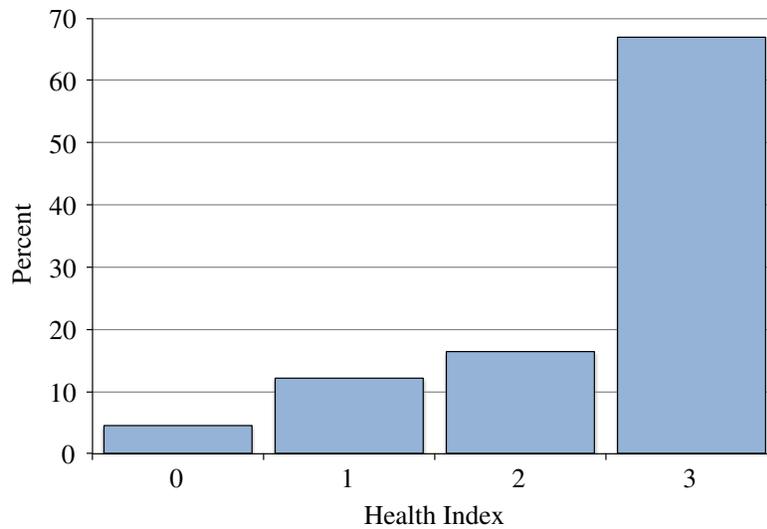
Fig. 2. *Baseline Respondent Health Index*
Note. Colour figure can be viewed at wileyonlinelibrary.com.
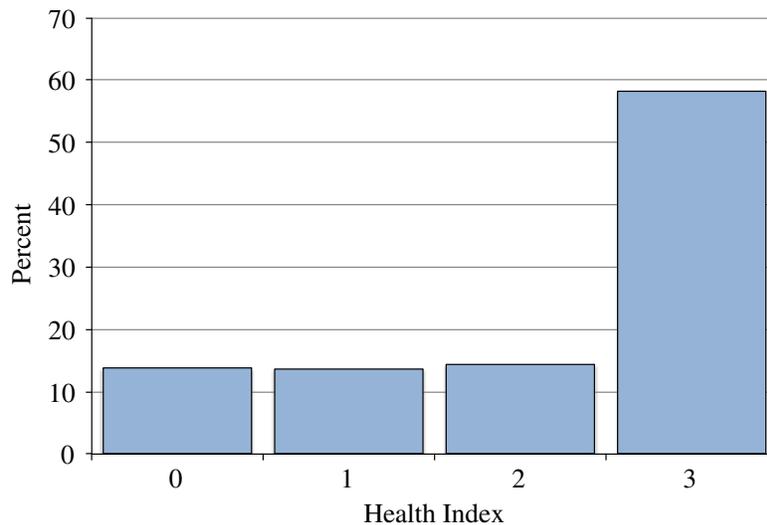
Fig. 3. *Baseline Child Health Index*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

We construct a traditional belief index (TBI) to measure adherence to Unani medicine. Under our main definition, the TBI is the unweighted sum of four binary variables, including whether the respondent believes:

(*i*) eating hot foods causes diarrhoea;
(*ii*) eating cold foods causes diarrhoea;
(*iii*) withholding foods is an effective diarrhoea treatment; and
(*iv*) withholding breast milk is an effective diarrhoea treatment.

The first two variables are core tenets of the belief system while the last two variables are Unani-based practices that follow from the belief in humoural balance. Figure 4 shows the baseline frequency distribution of the TBI, which equals zero or one for 43%

Fig. 4. *Baseline Traditional Belief Index*
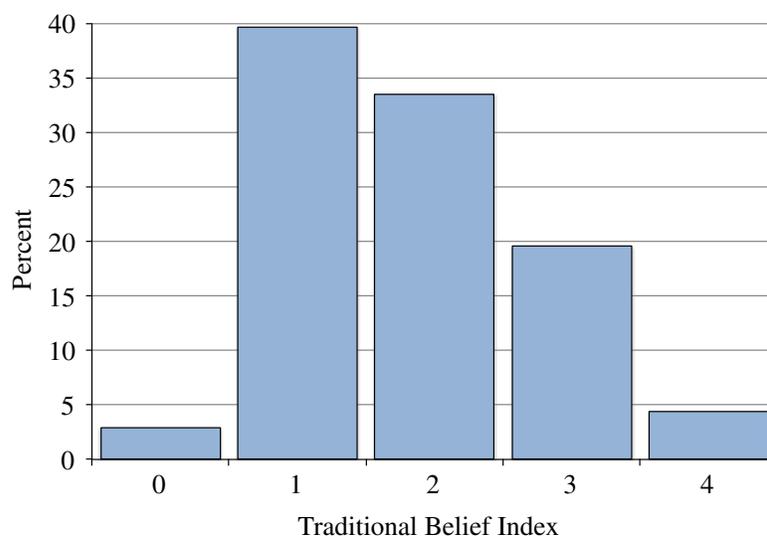*Note.* Colour figure can be viewed at wileyonlinelibrary.com.

of the sample. For specifications that interact traditional beliefs with treatment, we divide the sample into respondents for whom TBI ≤ 1 and TBI > 1.[12]

Our analysis includes three alternative constructions of the traditional belief index. The first alternative index is the first principal component of these elements. The second alternative is a 'summary index', which maximises the amount of information that is captured (Anderson, 2008). We prefer the unweighted sum because it is easier to interpret than either of these alternatives. Finally, we construct a broad index that includes the variables above, as well as whether the respondent:

  (v) has consulted a hakim (a Unani medical practitioner) in the past three months;
  (vi) would consult a hakim if her child had seizures; and
  (vii) would consult a hakim if her child were fainting.

We exclude these variables from the main index because they may reflect budget constraints and health, rather than beliefs. To facilitate interpretation, we normalise the alternative indices so that they have the same mean and standard deviation as the main index above. The baseline correlation between the main index and these three alternatives is 0.87, 0.95 and 0.85 respectively.

Our measure of traditional beliefs may also capture whether the respondent is informed or uninformed about hygiene. Theoretically, a lack of hygiene knowledge should strengthen treatment effects among high-TBI respondents, which is the

---

[12] The TBI may proxy for ignorance of modern medical practices as well as traditional medical beliefs. This relationship should encourage a larger impact of ML on high-TBI respondents since they have more scope to learn. Estimates below show the opposite pattern and do not support this interpretation. In addition, results are robust if we control for the interaction of treatment with ALC test scores, which proxy for cognitive ability.

opposite of our finding below. We develop a knowledge index (KI) to address this confound further. Respondents were asked to agree or disagree with the following four statements, all of which are false: (Q1) 'I can tell if my hands are clean just by looking at them'; (Q2) 'Untreated water is safe to drink'; (Q3) 'It is safe to eat food that has been touched by flies'; and (Q4) 'The worst thing diarrhoea can do is make a child uncomfortable'. We record whether the respondent answered correctly, as well as the total number of correct responses.[13] Unlike the elements of the TBI, the elements of the KI do not derive from traditional medicine. Second, the ML curriculum explicitly covers all components of the KI. The baseline correlation between the TBI and the KI is −0.02, which suggests that traditional medical beliefs are not necessarily incompatible with awareness of hygiene.

We assess intra-household informational spillovers with an endline sample of 2,057 female members of ALC participants' households. These respondents have a median age of 29 and 54% are married. Members of this sample are highly connected to the study participants. They say they are acquainted with 90% of nearby study participants and converse with 30% of nearby study participants at least weekly about health matters. Because of budget constraints, the data include self-reported hygiene and health for respondents and children but not weight-for-age and height-for-age or the appearance of the respondent's feet or clothes. Without baseline data, we cannot reproduce traditional belief interactions for this sample. Nobody in this sample participated in the intervention formally, however 27% indicate that they participated informally or observed in some way. To focus on social learning, our regressions distinguish between the 'full sample' (i.e. everyone who we interviewed) and the 'unexposed subsample,' who indicate that they did not participate (even informally) in the intervention.

Table 1 assesses the baseline balance of the estimation sample. Columns (1) to (3) show baseline means and columns (4) and (5) show p-values for the difference between the ML and IO arms and between the ML and C arms. By showing the estimation sample, the Table incorporates possible selection due to attrition and missing data, which we discuss in detail in Section A3 in the online Appendix.[14] We report 20 demographic and economic variables, including literacy, schooling, marital status, religious adherence, and assets. The Table does not show any imbalances in these variables that would pose a serious threat to validity. Table 1 also shows the baseline values of the dependent variables in our analysis. Hygiene, respondent health and traditional medical beliefs are balanced across treatment arms, however the knowledge index, the cleanliness of the cooking area and the presence of soap are higher in the IO arm (Table A9 shows that estimates are robust if we control for baseline covariates.)

---

[13] The knowledge index is problematic as an outcome variable for two reasons. First, Q1 and Q2 are imbalanced in the baseline, with ML respondents scoring significantly worse. Regressions that do not adequately address this imbalance may be biased downward. Secondly, Q1 and Q4 exhibit a downward time trend, suggesting that respondents become less informed over time. This pattern could reflect seasonality in the interpretation of these questions. For instance, respondents might answer Q4 differently during the monsoon period (when we conducted the endline survey) if they perceive that diarrhoea is more harmful at that time. Section A8 in the online Appendix shows the treatment effects and TBI interactions on this outcome. These results have the expected signs and are statistically significant.

[14] To compute p-values, we regress each variable on treatment and strata dummies and cluster by randomisation group. A similar Table for the full baseline sample is available from the authors and closely resembles Table 1.

Table 1

Baseline Characteristics for the Endline Estimation Sample by Treatment Status

	Mean			p-value	
	ML (1)	IO (2)	C (3)	ML − IO (4)	ML − C (5)
Demographic characteristics					
Baseline age	26.3	26.2	26.9	0.91	0.80
Illiterate	0.20	0.24	0.23	0.36	0.83
Any schooling	0.09	0.14	0.07	0.17	0.40
Married	0.57	0.55	0.58	0.69	0.83
Household size	7.0	7.0	7.2	0.99	0.97
Barailvi sect	0.87	0.86	0.83	0.66	0.46
Ramadan fasting days	12.1	12.8	11.1	0.42	0.31
Prays at least once per day	0.63	0.71	0.70	0.07*	0.17
Std. test score	−0.04	0.06	0.06	0.53	0.46
Std. change in test score	−0.09	−0.05	0.20	0.86	0.02**
Has children	0.32	0.30	0.31	0.66	0.45
Economic characteristics					
Improved roof	0.88	0.86	0.84	0.92	0.35
Bedrooms	2.3	2.2	2.3	0.35	0.80
Any savings	0.12	0.11	0.13	0.78	0.99
Land (acres)	3.2	4.3	2.6	0.39	0.33
Animals	0.66	0.71	0.70	0.47	0.26
Works outside the home	0.28	0.36	0.37	0.29	0.01***
Electricity	0.92	0.96	0.94	0.16	0.62
Refrigerator	0.29	0.30	0.25	0.89	0.22
Mobile phone	0.88	0.83	0.87	0.10*	0.62
Agriculture	0.51	0.51	0.49	0.94	0.82
Outcomes					
Hygiene (respondent)	2.36	2.35	2.34	0.75	0.34
Hygiene (children)	2.08	2.17	2.10	0.36	0.40
Lack of open defecation	2.00	1.97	2.01	0.82	0.60
Lack of garbage	1.98	1.92	1.98	0.50	0.76
Kitchen cleanliness	2.13	2.22	2.15	0.03**	0.63
Presence of soap	0.19	0.26	0.20	0.03**	0.97
Respondent health index	2.49	2.44	2.48	0.71	0.54
Child health index	2.15	2.03	1.95	0.16	0.20
Knowledge index	1.95	2.13	2.19	0.16	0.02**
Traditional belief index	1.80	1.79	1.89	0.93	0.79
Observations	1,144	1,099	979	–	–

Notes. p-values are based on OLS regressions with clustered standard errors that control for strata dummies. The sample is limited to endline non-attriters and matches the endline estimation sample. $N = 996$ for child hygiene, which has one observation per respondent with in-sample children. $N = 1,664$ for child health. †$p < 0.15$, *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

3. Estimation

3.1. *Specification*

Our regressions are based on a cross-sectional comparison of the treatment arms in the midline and endline rounds. Our primary regression specification yields separate midline and endline impacts:

$$Y_{ijt} = \beta_1(M_t \times ML_j) + \beta_2(E_t \times ML_j) + \beta_3(M_t \times IO_j)$$
$$+ \beta_4(E_t \times IO_j) + \beta_5 E_t + S_j + \varepsilon_{ijt}. \tag{1}$$

In this equation, i indexes the respondent and j indexes the ALC. ML and IO are indicators for the ML and Instruction Only arms. M indicates the three-month midline follow-up round and E indicates the 16-month endline follow-up round. Regressions do not include baseline data. Controlling for baseline characteristics does not appreciably change the estimates. Consistent with the stratified randomisation, regressions also control for strata dummies (S_j) (Kernan *et al.*, 1999). We cluster standard errors by randomisation group, allowing for arbitrary error correlations within groups, including within ALCs and households. We present intent-to-treat estimates throughout the article.

3.2. *Multiple Hypothesis Testing*

We consider the impact of ML and IO on several hygiene, sanitation, health and traditional medical belief outcomes. Standard single-hypothesis p-values are biased toward zero under multiple hypothesis testing because the inclusion of additional hypotheses increases the probability of a false discovery (Romano *et al.*, 2008). We adjust p-values and significance levels for multiple testing with the Romano and Wolf (2005) stepdown procedure. This procedure controls for the family-wise error rate, which is the probability of one or more false discoveries. The stepdown is more powerful than traditional multiple testing corrections like Bonferroni because it accounts for the correlation in statistical significance across hypotheses. Empirical articles by Heckman *et al.* (2010, 2013), Lee and Shaikh (2014) and Augsburg *et al.* (2015) also use this approach. The use of indices, such as the health index and traditional belief index in our analysis, is another way to address multiple testing (Katz *et al.*, 2001; Anderson, 2008).

We adjust for multiple testing across outcomes within groups of related hypotheses. In principle, every coefficient and comparison of coefficients is a potential hypothesis. However, family-wise error control is unrealistically conservative for large sets of hypotheses. We follow Heckman *et al.* (2010) and group hypotheses by outcome type and regressor. The authors note that 'there is some arbitrariness in defining the blocks of hypotheses that are tested in a multiple-testing procedure' (p. 24). The choice of groups affects the interpretation but not the internal validity of the adjusted p-values. Our approach addresses the use of multiple outcome variables, which is a primary multiple testing concern. It does not address the multiplicity of hypotheses within any regression with more than one independent variable. We also leave aside the issue of multiple testing across robustness estimates, such as regressions with additional controls. We also do not adjust for the presence of multiple outcome groups.

We classify outcomes into six categories: respondent hygiene, child hygiene and household sanitation, health and anthropometrics, hygiene of other household members, health of other household members, and traditional medical beliefs. This classification leads to Table × row hypothesis groups for our main treatment effect estimates in Tables 2–4. The note below each regression Table clarifies precisely how

we group hypotheses and compute p-values.[15] Stars indicate multiple-testing-adjusted significance levels unless otherwise indicated, however the reader can compare coefficients to standard errors to determine unadjusted significance levels. Multiple testing adjustments based on other hypothesis groupings are available from the authors.[16]

3.3. *Results*

Table 2 shows estimates for respondent hygiene. In column (1), ML improves hygiene by 0.14 points on a 3-point scale (0.25σ) in the three-month midline survey and by 0.16 points (0.27σ) in the 16-month endline survey. In contrast, IO has a statistically-insignificant impact of 0.01 in both rounds. The difference between the impacts of ML and IO is also statistically significant in both rounds. Columns (2)–(5) show impacts of ML and IO on the cleanliness of hands, fingernails, feet, and clothing in the endline survey. Assessments of hands, fingernails and feet are recorded on 3-point scales while the clothing assessment is recorded on a 4-point scale. As with overall hygiene, the impact of ML on these outcomes is strong and significant (from 0.23σ to 0.30σ) while the impact of IO is small and insignificant.[17] For comparison, the baseline correlational effect on hygiene of an additional year of schooling is 0.02 points and the difference between literate and illiterate respondents is 0.12 points.

Table 3 shows estimates for child hygiene and household sanitation. These outcomes are household public goods that depend in part on the actions of other household members. In column (1), ML improves child hygiene by 0.07 points on a 3-point scale (0.11σ) in the midline survey and by 0.18 points (0.29σ) in the endline survey. The endline result is robust to the multiple testing adjustment while the midline result is not. The impact of IO is smaller and is not significant in either round. Columns (2)–(5) show household sanitation estimates. Neither ML or IO improves household sanitation in the midline survey. However, ML improves three outcomes in the endline survey. The absence of garbage improves by 0.18 points on a 3-point scale (0.27σ), the cleanliness of the cooking area improves by 0.12 points on a 3-point scale (0.25σ) and probability of observing soap increases by 10 percentage points. Adjusted p-values are less than 0.10 for each of these estimates.

The persistence and strengthening of these impacts over time may be surprising if people subsequently receive other signals or forget the intervention message. However, there are several channels through which effects may strengthen for both individuals

---

[15] Table 6 shows that the traditional belief interaction is robust across several alternative specifications. We do not adjust for multiple testing here because the regressions have a common dependent variable. Table 8 provides treatment effect estimates for four alternative definitions of the TBI, as well as the four components of the main TBI. We group the index versions and the components separately. Combining these outcomes into one group would address multiple testing twice since the use of an index is itself an appropriate response to multiple testing.

[16] Hygiene and sanitation estimates are robust if we treat hygiene and sanitation as a single category. The midline and endline impacts of ML on respondent hygiene have p-values of 0.10 and 0.07 in this case.

[17] Section A8 in the online Appendix shows a similar pattern of results for hygiene knowledge. ML significantly increases the belief that clean hands may still be contaminated and that untreated water may be unsafe. Effects are weaker for two other items that are tangential to the role of microbes, which suggests that the intervention works by raising awareness of microbes.

Table 2

The Impact on Respondent Hygiene

| | Respondent hygiene | | | | |
	Overall (1)	Hands (2)	Nails (3)	Feet (4)	Clothes (5)
Midline × Microbe Literacy	0.14** (0.064)	–	–	–	–
Midline × Instruction Only	0.022 (0.064)	–	–	–	–
Endline × Microbe Literacy	0.16** (0.062)	0.29** (0.10)	0.32** (0.11)	0.29** (0.11)	0.14** (0.059)
Endline × Instruction Only	0.013 (0.060)	−0.060 (0.100)	−0.063 (0.11)	−0.056 (0.11)	−0.024 (0.056)
p-value: ML − IO (mid)	0.07	–	–	–	–
p-value: ML − IO (end)	0.02	0.01	0.01	0.01	0.02
Dependent variable mean (C)	2.27	2.97	2.81	2.78	2.34
Observations	7,103	3,327	3,327	3,327	3,327

Notes. Clustered standard errors appear in parentheses. The Table reports multiple-testing-adjusted p-values and significance levels. Hypotheses are grouped by row, as we describe in the text. Regressions control for strata and round dummies. †$p < 0.15$, *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 3

The Impact on Child Hygiene and Household Sanitation

	Child hygiene (1)	Lack of defecation (2)	Lack of garbage (3)	Clean cooking area (4)	Soap present (5)
Midline × Microbe Literacy	0.069 (0.070)	0.057 (0.095)	−0.0051 (0.066)	−0.026 (0.048)	−0.025 (0.042)
Midline × Instruction Only	0.11 (0.062)	−0.024 (0.097)	0.025 (0.072)	0.025 (0.042)	0.041 (0.039)
Endline × Microbe Literacy	0.18* (0.067)	0.14 (0.11)	0.21** (0.088)	0.13** (0.050)	0.10* (0.043)
Endline × Instruction Only	0.082 (0.063)	0.014 (0.11)	0.0072 (0.080)	0.067 (0.054)	0.12* (0.044)
p-value: ML − IO (mid)	0.85	0.79	0.67	0.62	0.35
p-value: ML − IO (end)	0.42	0.38	0.07	0.34	0.74
Dependent variable mean (C)	1.87	2.05	2.02	2.15	0.23
Observations	2,028	7,103	7,103	7,103	7,103

Notes. Clustered standard errors appear in parentheses. The Table reports multiple-testing-adjusted p-values and significance levels. Hypotheses are grouped by row, as we describe in the text. All regressions control for strata dummies and round dummies. †$p < 0.15$, *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

and households. At the individual level, behaviour change may foster a virtuous cycle of learning-by-doing in which health improvements reinforce the validity of the hygiene message. Results are also consistent with habit formation. In addition, households may face frictions related to social learning, bargaining, or fixed investments, that limit the short-term impact of information. Because of these multiple channels, it is unclear whether we should expect effects to strengthen or weaken over time.

Table 4

The Impact on Respondent and Child Health

	Respondent	Children		
	Health index (1)	Health index (2)	WAZ (3)	HAZ (4)
Midline × Microbe Literacy	0.21***	0.091	–	–
	(0.069)	(0.080)		
Midline × Instruction Only	0.063	−0.031	–	–
	(0.072)	(0.077)		
Endline × Microbe Literacy	0.18***	0.050	0.27	0.29
	(0.063)	(0.088)	(0.18)	(0.17)
Endline × Instruction Only	0.056	0.079	0.083	0.19
	(0.061)	(0.083)	(0.19)	(0.17)
p-value: ML−IO (mid)	0.06	0.16	–	–
p-value: ML−IO (end)	0.24	0.75	0.33	0.58
Dependent variable mean (C)	2.38	2.28	−1.90	−2.45
Observations	7,103	3,908	1,395	1,395

Notes. Clustered standard errors appear in parentheses. The Table reports multiple-testing-adjusted p-values and significance levels. Hypotheses are grouped by row, as we describe in the text. All regressions control for strata and round dummies. †p < 0.15, *p < 0.1, **p < 0.05, ***p < 0.01.

Results for the health index and child weight and height appear in Table 4. In column (1), ML increases the health index (reduces the number of morbidities) for respondents by 0.21 points (0.28σ) at midline and by 0.18 points (0.22σ) at endline. The impact of IO is around 0.06 and is statistically insignificant. Although studies often focus directly on child morbidity, adult health may have important intergenerational spillovers, particularly since many study women are likely to become pregnant in the immediate future. Column (2) shows weaker effects on the health index for children. ML increases the index by 0.08 points (0.09σ) at midline and by 0.07 points (0.05σ) at endline. Both results are insignificant regardless of the multiple testing adjustment. The clearer results for other outcomes suggest that measurement error may interfere with the child health estimates. We suspect that respondents may systematically underreport child morbidity because they are unaware of or have forgotten some instances of morbidity (Zafar *et al.*, 2010; Lamberti *et al.*, 2015). This phenomenon works to reduce the difference between treatment arms and bias treatment effect estimates toward zero. We do not find significant differences between the health impacts for boys and girls. Results for disaggregated morbidities appear in Section A6 in the online Appendix.

Columns (3) and (4) of Table 4 show impacts on child weight and height.[18] WAZ and HAZ are 0.27 and 0.29 points higher, respectively, in ML than in C. These differences are not statistically significant: unadjusted p-values are around 0.10 and

[18] We cannot validate pre-intervention balance for weight and height because these outcomes are only present at endline. One possible concern is that baseline child health is somewhat higher (although not significantly different) in ML than in IO or C in Table 1. To assess the severity of this issue, we regress WAZ and HAZ on the health index for control children and multiply the coefficients from these regressions by the baseline health difference between ML and C. This exercise suggests that the baseline health imbalance explains 8% of the WAZ estimate and 13% of the HAZ estimate in Table 4. In addition, WAZ and HAZ estimates that control for baseline health closely resemble the estimates in the article.

adjusted p-values are around 0.20. However, we show below that there is considerable heterogeneity in this impact, so that the entire effect occurs for children of respondents with weak traditional medical beliefs. The impact on weight closely resembles Ahmed *et al.*'s (1994) estimate 0.28 for a community-based intervention featuring information about the germ theory. ML and C children differ in weight by around 0.86 kilograms, which is roughly equivalent to 6,622 calories (Wishnofsky, 1958). To achieve this difference over 16 months, ML children would need to absorb around 14 additional calories per day.[19]

The hygiene and health impacts of ML should covary positively if hygiene has a positive marginal product. However, the reduced-form results above do not clarify whether the same people who realise better hygiene also realise better health. To consider this prediction, we jointly bootstrap the impacts of ML on respondent hygiene and health. This approach yields the correlation between the hygiene and health treatment effects across bootstrap replications. Figure 5 shows a scatterplot of replication ML coefficients and for respondent hygiene on the *x*-axis and respondent health on the *y*-axis. The distinct positive relationship in the Figure indicates that replications with strong hygiene impacts also tend to have strong health impacts. Section A5 in the online Appendix shows the correlation between all behaviour and health outcomes in our analysis. The correlation coefficient of 0.41 between respondent hygiene and health impacts is higher than for any sanitation outcome. The impact on respondent hygiene is also strongly correlated with the impact on child weight and height. These results are consistent with a positive marginal product of hygiene, however they do not rule out the possibility of a health impact though an unobservable channel that is correlated with hygiene.[20]

Finally, Table 5 shows the impact of ML and IO on the hygiene and health of other adult women in study participants' households. Panel (*a*) shows estimates for the full sample and panel (*b*) shows estimates for the unexposed subsample, who say they did not participate in the intervention. In panel (*a*), ML improves overall hygiene by 0.21 points (0.31σ) and the cleanliness of both hands and fingernails by 0.31 points (0.30σ for both outcomes). We use seemingly unrelated regression to compare the magnitudes of these effects to our main estimates above. Although estimates for other household members are larger, the differences are not statistically significant for any outcome.[21] In column (4), ML improves respondent health by 0.25 points (0.27σ). In

---

[19] Water supply, sanitation, and hygiene interventions have heterogeneous anthropometric effects in the literature (Fewtrell *et al.*, 2005). Fenn *et al.* (2012) show that an intervention combining water source protection and hygiene and sanitation education improved the HAZ by 0.33 points among children under 3 in rural Ethiopia. However, many studies do not find effects on these outcomes (Kremer *et al.*, 2011; Bowen *et al.*, 2012). Economic shocks also appear to have significant anthropometric impacts. Galiani and Schargrodsky (2004) find that land titling improves the WAZ by 0.2 to 0.3 points in Argentina. Hidrobo (2014) shows that exposure to an economic crisis in Ecuador reduces the HAZ by 0.08 points.

[20] Other ways of exploring the causal chain do not work well in this setting. As a proxy variable, personal appearance may explain relatively little of the variation in health in a Blinder–Oaxaca decomposition. A strategy to estimate the marginal product of hygiene cross-sectionally (as in a Mincer regression) is unlikely to be informative because households simultaneously optimise over substitutable health inputs.

[21] It is difficult to compare treatment effects for study participants and other household members because these groups received fundamentally different treatments. Health information from a family member may be inherently more credible than information from an outside source. These groups also have different demographic characteristics. Other household members are older and have fewer young children, which may facilitate compliance.
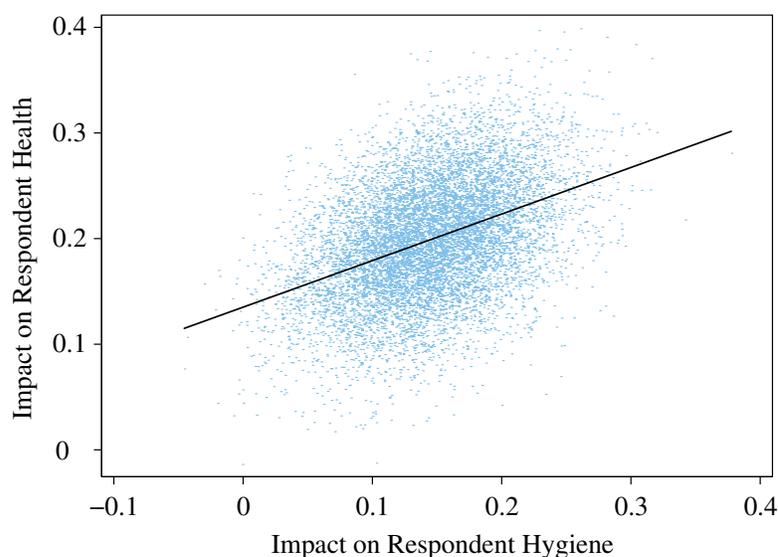
Fig. 5. *The Impact of Microbe Literacy on Hygiene and Health Across Bootstrap Replications*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

Table 5

The Impact on Other Female Household Members

	Hygiene			Health index
	Overall (1)	Hands (2)	Nails (3)	(4)
Panel (*a*): full sample				
Endline × Microbe Literacy	0.21***	0.31**	0.31**	0.25***
	(0.061)	(0.11)	(0.11)	(0.082)
Endline × Instruction Only	0.039	0.022	0.058	0.072
	(0.069)	(0.12)	(0.12)	(0.090)
p-value: ML − IO	0.04	0.03	0.03	0.05
p-value: other female − main sample (ML)	0.38	0.76	0.87	0.37
Dependent variable mean (C)	2.43	3.10	3.02	2.40
Observations	2,057	2,057	2,057	2,057
Panel (*b*): unexposed subsample				
Endline × Microbe Literacy	0.15*	0.21*	0.22†	0.13
	(0.068)	(0.12)	(0.12)	(0.081)
Endline × Instruction Only	0.033	−0.012	0.022	0.067
	(0.070)	(0.12)	(0.12)	(0.084)
p-value: ML − IO	0.11	0.14	0.16	0.33
p-value: other female − main sample (ML)	0.83	0.35	0.22	0.62
Dependent variable mean (C)	2.39	3.05	2.96	2.39
Observations	1,497	1,497	1,497	1,497

Notes. Clustered standard errors appear in parentheses. p-values and significance levels are adjusted for multiple hypothesis testing by row and outcome category. All regressions control for strata dummies. The unexposed subsample consists of people who did not participate in either intervention component. 'Other female − main sample' p-values use seemingly unrelated regression to test whether estimates for other household members differ from the main sample estimates in Tables 2 and 4. †$p < 0.15$, *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

© 2018 Royal Economic Society.

contrast, the effects of IO are generally small and insignificant. Panel (*b*) shows that effects are weaker but qualitatively similar for the unexposed sample. These estimates represent a lower bound on the impact via information spillovers since many 'exposed' household members likely did not participate extensively.

3.4. *Interpretation*

The estimates in Tables 2–5 show that the microscope demonstration significantly magnifies the impact of hygiene information. Here we interpret this result through a simple Bayesian learning model, which we develop formally in Section A1 in the online Appendix. In principle, the microscope demonstration may enhance either the mean or the precision of the informational signal. Advocates of the programme believe that ML increases the signal precision because a demonstration of the existence of microbes makes the infection-prevention message more credible. However, the microscope demonstration could also increase the signal precision in other ways. For instance, the microscope is a novel object that could enhance the credibility or prestige of the facilitators. ML could also increase the mean of the signal. ML involves twice as much instructional time as IO. Although the microscope demonstration curriculum officially omits hygiene instruction, facilitators might provide information informally, such as in response to questions. The practice of obtaining slide samples from 'dirty' substances like standing water could indirectly convey a negative message about microbes.

Despite these possibilities, evidence suggests that ML works by increasing the awareness of microbes. The microscope demonstration precedes the infection-prevention workshop by several days, making it unlikely that the spectacle of the demonstration simply makes participants more alert. Field reports also suggest that participants responded to the visualisation of microbes. One facilitator remarked, 'Before the workshop, people did not accept that microbes exist. But when we collected samples and made the slides, then they believed'. Another facilitator commented, 'Before the intervention, learners said they didn't know about germs. When we showed them the microbes on the screen, they were amazed'.

Evidence of the impact on hygiene knowledge also speaks to this point. Section A8 in the online Appendix shows treatment effects on hygiene knowledge. We asked respondents to agree or disagree with four factual statements, all of which are false: (Q1) 'I can tell if my hands are clean just by looking at them'; (Q2) 'Untreated water is safe to drink'; (Q3) 'It is safe to eat food that has been touched by flies'; and (Q4) 'The worst thing diarrhoea can do is make a child uncomfortable'. In Table A12, ML has the strongest effect on Q1 and Q2, which relate most directly to microbes and has weaker effects on Q3 and Q4. Moreover, Table A8 shows that the correlation in the impacts on knowledge and respondent health across bootstrap replications is 0.42.

4. The Role of Traditional Medicine

This Section examines the role of traditional medical beliefs and the impact on hygiene learning. We interact treatment status with the baseline TBI to examine whether traditional beliefs moderate the impact of information. Then we estimate the impact of the intervention on the TBI. The model in Section A1 in the online Appendix shows that

traditional medical beliefs moderate the impact of information in an ambiguous way. Traditional beliefs may interfere with learning by giving people more precise priors about the effectiveness of hygiene. However, they may also give people lower-mean priors, which create more scope for learning. In the model, a negatively signed interaction between hygiene education and traditional medical beliefs is evidence that traditional beliefs operate on the precision of priors. Traditional medical beliefs may be correlated with other factors that influence learning, such as socio-economic status, cognitive ability, and prior hygiene and health experience. These omitted factors threaten to confound the interaction between treatment and traditional medical beliefs. We address the potential for omitted variables by controlling for the interaction between treatment and 34 demographic, economic and academic covariates.[22]

Table 6 shows the differential effects of ML and IO on hygiene for believers in traditional medicine. Since the intervention has similar midline and endline hygiene effects, we pool the midline and endline rounds to increase power. Our primary specification in column (1) uses the main TBI definition and distinguishes between low-TBI and high-TBI respondents as above. The ML coefficient indicates that ML improves the hygiene of low-TBI respondents by 0.23 points. The sum of the ML and $ML \times TBI_H$ coefficients indicates that ML only improves the hygiene of high-TBI respondents by 0.08 points, a significant difference from the TBI_L response. Both the IO and $IO \times TBI_H$ coefficients are small and insignificant. Columns (2)–(4) of Table 6 show the TBI_H interaction under alternative TBI definitions. Column (2) uses the principal component version, column (3) uses the summary index version, and column (4) uses the broad version of the TBI. We define TBI_H to capture the top 60% of the TBI distribution in each case, so that estimates have a comparable interpretation to column (1). Effects are remarkably similar across these specifications. The $ML \times TBI_H$ coefficient is statistically significant and ranges from -0.13 to -0.15 while the $IO \times TBI_H$ coefficient is uniformly small and insignificant. These results suggest that the interaction with traditional medical beliefs is not sensitive to the particular TBI definition.

Columns (5)–(8) of Table 6 assess whether omitted variables confound the TBI_H interaction estimate. Each specification controls for the interaction between ML and IO and multiple baseline covariates, which should attenuate the $ML \times TBI_H$ coefficient if the result is spurious. We demean the covariates within the TBI_L group so that the main effects of ML and IO remain comparable to columns (1)–(4). Column (5) includes interactions with the demographic and economic variables in Table 1 except ALC test scores. Column (6) includes interactions with baseline hygiene, sanitation and respondent and child health. Column (7) includes the interaction with ALC test scores and column (8) includes the interactions with all of these variables at once. While the controls are highly jointly significant in all specifications, the

---

[22] We assess the potential for omitted variables bias in Table A13 by comparing the baseline characteristics of low-TBI and high-TBI respondents (TBI_L and TBI_H). Despite this concern, traditional beliefs are not strongly correlated with other baseline characteristics. Both groups have similar hygiene and health patterns. Low-TBI respondents are younger and more literate, although these differences are insignificant. There is no significant difference across groups in the level or change in ALC test scores, which suggests that these groups have comparable cognitive ability. In addition, assignment to treatment, treatment compliance and attrition are uncorrelated with traditional beliefs.

Table 6

Traditional Medical Beliefs and the Impact on Respondent Hygiene

	Respondent hygiene							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Microbe Literacy	0.25***	0.24***	0.25***	0.25***	0.22***	0.22***	0.20**	0.20***
	(0.078)	(0.080)	(0.077)	(0.079)	(0.073)	(0.075)	(0.080)	(0.067)
Microbe Literacy × TBI$_H$	−0.17**	−0.15**	−0.15**	−0.15**	−0.13**	−0.14**	−0.15**	−0.10*
	(0.073)	(0.075)	(0.073)	(0.073)	(0.065)	(0.065)	(0.070)	(0.058)
Instruction Only	0.0082	−0.0014	0.0046	−0.0043	−0.0058	0.011	−0.038	0.016
	(0.063)	(0.066)	(0.063)	(0.065)	(0.075)	(0.069)	(0.070)	(0.064)
Instruction Only × TBI$_H$	0.025	0.039	0.031	0.043	0.0072	0.015	0.045	0.019
	(0.061)	(0.063)	(0.060)	(0.061)	(0.061)	(0.058)	(0.059)	(0.056)
p-value: ML & IO × covariates	–	–	–	–	0.00	0.00	0.00	0.00
TBI definition	Main	PC	Summary	Broad	Main	Main	Main	Main
Demo and economic controls	–	–	–	–	Yes	–	–	Yes
Hygiene and health controls	–	–	–	–	–	Yes	–	Yes
Literacy test score controls	–	–	–	–	–	–	Yes	Yes
Dependent variable mean (C)	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27
Observations	7,103	7,103	7,103	7,103	7,103	7,103	7,103	7,103
R²	0.04	0.04	0.04	0.04	0.09	0.12	0.06	0.16

Notes. Clustered standard errors appear in parentheses. All regressions control for strata and round dummies. The Table reports unadjusted p-values and significance levels. †p < 0.15, *p < 0.1, **p < 0.05, ***p < 0.01.

ML $\times$ TBI$_H$ coefficient is generally insensitive to these controls. The estimate is attenuated by 7–27% in columns (5)–(7) and by 35% in column (8) but remains statistically significant. These estimates suggest that the TBI$_H$ interaction does not arise through a spurious correlation with other determinants of learning.[23]

TBI$_H$ interactions for health and anthropometric outcomes (following the specification in column (1) of Table 6) appear in Table 7. In columns (1) and (2), the differential effects on the respondent and child health indices have the expected sign but are not statistically significant. Columns (3) and (4) show results for child weight and height. ML increases the WAZ and HAZ by 1.01 and 1.09 points, respectively, for children of low-TBI respondents. It has no impact on these outcomes for children of high-TBI respondents. These estimates are robust to multiple hypothesis testing across outcomes. Estimates with alternative TBI definitions and additional controls (analogous to columns (2)–(8) of Table 6) appear in Section A9 in the online Appendix.

Table 8 shows the impact of the intervention on traditional medical beliefs. While it did not directly discourage Unani medicine, ML could have weakened traditional beliefs by substantiating the germ theory. Columns (1)–(4) show impacts on the four alternative TBI versions while columns (5)–(8) show impacts on the four components of the main TBI.[24] Column (1) shows that in the midline survey ML reduces the TBI by 0.16 points (0.24σ) while IO reduces the TBI by 0.08 points (0.12σ, an insignificant result). Columns (2)–(4) show a similar pattern for alternative TBI definitions. In principle, this effect could arise through either a leftward shift in the distribution or a reallocation of mass to the left tail. The midline TBI frequency distributions for ML and C (available from the authors) show a shift in mass from TBI = 2 and 3 into TBI = 1 suggesting that the intervention led people to moderate but not abandon their traditional beliefs.

Columns (5)–(8) of Table 8 show estimates for the four components of the main TBI. In columns (5) and (6), the intervention does not change beliefs that hot and cold foods cause diarrhoea, which are core aspects of the Unani belief system. Instead, columns (7) and (8) show that ML reduces perceptions that withholding food and milk are appropriate diarrhoea treatments, which are implications of the Unani model. These perceptions fall by 6–8% in the midline survey. The impact on the TBI and its components does not persist in the endline survey. There is no significant difference across arms in the TBI or its components after 16 months. These results suggest that the intervention had only a modest and temporary impact on traditional beliefs.[25]

Section A9 in the online Appendix includes several additional robustness tests related to the TBI$_H$ interaction. Table A14 shows similar TBI interaction estimates for an alternative hygiene indicator. Table A15 provides TBI$_H$ interaction estimates for all

---

[23] TBI$_H$ interactions for child hygiene and household sanitation are statistically insignificant and are available from the authors. Conceptually, these regressions should incorporate the baseline traditional beliefs of other household members, which we do not observe. The ML $\times$ TBI$_H$ coefficient for respondent hygiene in column (1) of Table 6 has a p-value of 0.19 if we adjust for multiple testing across hygiene and sanitation outcomes.

[24] We separately adjust for multiple hypothesis testing across TBI versions and TBI components, although the use of an index already addresses the multiplicity of traditional belief components.

[25] This result contrasts with the lasting hygiene and health impacts above, and suggests that traditional beliefs may interfere with learning related to modern medicine but not with adherence to practices people have already learned. More research is needed to explore these interactions further.

Table 7

Traditional Medical Beliefs and the Impact on Health

	Respondent Health (1)	Children		
		Health (2)	WAZ (3)	HAZ (4)
Microbe Literacy	0.25**	0.13	1.03***	1.12***
	(0.086)	(0.11)	(0.30)	(0.30)
Microbe Literacy $\times$ TBI$_H$	-0.099	-0.11	-1.10***	-1.22***
	(0.082)	(0.12)	(0.30)	(0.31)
Instruction Only	0.033	0.0085	0.61*	0.51*
	(0.080)	(0.10)	(0.30)	(0.28)
Instruction Only $\times$ TBI$_H$	0.052	0.033	$-0.77^\dagger$	-0.46
	(0.077)	(0.099)	(0.34)	(0.33)
Dependent variable mean (C)	2.38	2.28	-1.90	-2.45
Observations	7,103	3,908	1,395	1,395

Notes. Clustered standard errors appear in parentheses. The Table reports multiple-testing-adjusted p-values and significance levels. Hypotheses are grouped by row, as we describe in the text. All regressions control for strata and round dummies. $\dagger p < 0.15$, $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

health outcomes while controlling for the interaction of ML and IO with baseline covariates (analogous to columns (5)–(8) of Table 6). Finally, Table A17 provides evidence that traditional beliefs and the lack of hygiene knowledge are conceptually distinct. The Table shows that TBI$_H$ estimates are robust if we control for KI$_L$, an indicator for low baseline hygiene knowledge. The KI$_L$ interactions are positive but insignificant, which suggests that uninformed people respond more to information but that this channel is not particularly strong.

5. Conclusion

Public health messages must be convincing in order to change behaviour. Prescriptive messages about infectious disease prevention, which typically rely on the germ theory of disease, may be unpersuasive to people who are unfamiliar with microbes or believe in an alternative disease model. The strong and lasting impact of the microscope demonstration suggests that the unfamiliarity with microbes is an important barrier to hygiene adoption. While the design does not isolate awareness of microbes as the mechanism, the additional results and contextual factors in subsection 3.4 support this interpretation.

Training, transportation and wages are the main expenses of ML, which costs US$4.95 per participant as implemented (microscopes have a small rental cost because they last for many years).[26] The strong endline results for other household members suggest that

---

[26] The total implementation costs were PKR 1,001,200 (US$10,166 using an exchange rate of $1 = PKR 98.49 from 1 May 2013). We allocate 2/3 of these costs to ML based this arm's share of intervention contact time, and divide by the number of study participants in the ML arm (1,351). This cost breaks down into $2.80 for field work and $2.15 for training. Idiosyncratic aspects of the training for this study led to high training costs, and these costs could be substantially reduced in subsequent implementations of ML. Using the costs of individual programme components, MLI officials estimate that it would cost $2.31 per participant to implement a scaled-up ML intervention reaching 5,000 participants.

Table 8

The Impact on Traditional Medical Beliefs

	Traditional belief index				Diarrhoea causes		Diarrhoea treatments	
	Main (1)	PC (2)	Sum. (3)	Broad (4)	Hot foods (5)	Cold foods (6)	No food (7)	No milk (8)
Midline × Microbe Literacy	−0.16**	−0.20***	−0.13**	−0.11*	0.015	−0.041	−0.058†	−0.078*
	(0.058)	(0.055)	(0.055)	(0.058)	(0.020)	(0.030)	(0.029)	(0.032)
Midline × Instruction Only	−0.091	−0.10	−0.092	−0.060	0.0060	−0.041	−0.0029	−0.053
	(0.071)	(0.074)	(0.063)	(0.067)	(0.021)	(0.033)	(0.034)	(0.038)
Endline × Microbe Literacy	0.033	0.020	0.027	0.051	0.012	−0.016	0.017	0.017
	(0.069)	(0.066)	(0.066)	(0.067)	(0.018)	(0.051)	(0.035)	(0.026)
Endline × Instruction Only	0.020	0.047	0.010	−0.017	−0.0089	−0.019	0.0049	0.043
	(0.064)	(0.063)	(0.065)	(0.066)	(0.021)	(0.047)	(0.025)	(0.023)
p-value: ML − IO (mid)	0.41	0.33	0.57	0.52	0.82	0.98	0.24	0.86
p-value: ML − IO (end)	0.84	0.92	0.87	0.59	0.60	0.88	0.95	0.79
Dependent variable mean (C)	1.40	1.42	1.40	1.39	0.93	0.21	0.09	0.16
Observations	7,103	7,103	7,103	7,103	7,103	7,103	7,103	7,103

Notes. Clustered standard errors appear in parentheses. p-values and significance levels are adjusted for multiple hypothesis testing by row and outcome category. Columns (1)–(4) and columns (5)–(8) are distinct outcome categories for this procedure. All regressions control for strata and round dummies. †p < , *p < 0.1, **p < 0.05, ***p < 0.01.

information eventually disseminates to non-participants, increasing the intervention's impact and cost-effectiveness. In light of the large and persistent impact of this intervention, policymakers should consider ML in other settings. Offering individually tailored information about hand and water contamination and appealing directly to traditional medicine may be other creative ways to overcome scepticism about microbes.

More work is needed to understand the ramifications of traditional medicine. Our findings suggest that traditional medicine may discourage hygiene and foster infectious disease. A substantial share of the world population adheres to some version of traditional medicine, and traditional medicine may have important global health consequences if this pattern extrapolates to other settings. While traditional medicine is diverse and some forms may be consistent with healthy behaviour, it is puzzling that unambiguously harmful practices persist in equilibrium. The opacity of the health production function may contribute to this phenomenon, as placebo effects and mean reversion lead people to believe in erroneous causal effects. Hakims are cheaper and more available than Western doctors in this setting, and traditional beliefs may persist because pro-traditional messages are less costly than pro-modern messages.

University of Southern California
Vienna University of Economics and Business
London School of Hygiene and Tropical Medicine

Additional Supporting Information may be found in the online version of this article:

Appendix A. Additional Study Details and Results
Data S1.

References

Ahmad, F., Ahmad, U., Aziz, B., Higgins, E., King, I., Naqvi, A. and Rehman, I. (2012). 'Preliminary results on use of microscopes to reduce rates of diarrhea and respiratory illness in children under five of non-literate parents in Swat Valley, Pakistan', Unpublished manuscript.

Ahmed, N. and Zeitlan, M. (1993). 'Assessment of the effects of teaching germ theory on changes in hygiene behaviors, cleanliness, and diarrheal incidence in rural Bangladesh', *International Quarterly of Community Health Education*, vol. 14(3), pp. 283–97.

Ahmed, N., Zeitlin, M., Beiser, A., Super, C. and Gershoff, S. (1993). 'A longitudinal study of the impact of behavioural change intervention on cleanliness, diarrhoeal morbidity and growth of children in rural Bangladesh', *Social Science and Medicine*, vol. 37(2), pp. 159–71.

Ahmed, N., Zeitlin, M., Beiser, A., Super, C., Gershoff, S., and Ahmen, M. (1994). 'Assessment of the impact of a hygiene intervention on environmental sanitation, childhood diarrhoea, and the growth of children in rural Bangladesh', *Food and Nutrition Bulletin*, vol. 15(1), pp. 40–52.

Aiello, A., Coulborn, R., Perez, V. and Larson, E. (2008). 'Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis', *American Journal of Public Health*, vol. 98(8), pp. 1372–81.

Almond, D. (2006). 'Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population', *Journal of Political Economy*, vol. 114(4), pp. 672–712.

Anderson, M. (2008). 'Multiple inference and gender differences in the effects of early interventions: a reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American Statistical Association*, vol. 103(484), pp. 1481–95.

Anwar, M., Green, J. and Norris, P. (2012). 'Health-seeking behaviour in Pakistan: a narrative review of the existing literature', *Public Health*, vol. 126(6), pp. 507–17.

Argent, J., Augsburg, B. and Rasul, I. (2014). 'Livestock asset transfers with and without training: evidence from Rwanda', *Journal of Economic Behavior and Organization*, vol. 108, pp. 19–39.

Augsburg, B., De Haas, R., Harmgart, H. and Meghir, C. (2015). 'The impacts of microcredit: evidence from Bosnia and Herzegovina', *American Economic Journal: Applied Economics*, vol. 7(1), pp. 183–203.

Avitabile, C. (2012). 'Does information improve the health behavior of adults targeted by a conditional transfer program?', *Journal of Human Resources*, vol. 47(3), pp. 785–825.

Barnes, P., Bloom, B. and Nahin, R. (2008). 'Complementary and alternative medicine use among adults and children: United States, 2007', National Health Statistics Report No. 12.

Bowen, A., Agboatwalla, M., Luby, S., Tobery, T., Ayers, T. and Hoekstra, R.M. (2012). 'Association between intensive handwashing promotion and child development in Karachi, Pakistan', *Archives of Pediatrics and Adolescent Medicine*, vol. 166(11), pp. 1037–44.

Bratti, M. and Mendola, M. (2014). 'Parental health and child schooling', *Journal of Health Economics*, vol. 35, pp. 94–108.

Cairncross, S., Shordt, K., Zacharia, S. and Govindan, B.K. (2005). 'What causes sustainable changes in hygiene behaviour? A cross-sectional study from Kerala, India', *Social Science and Medicine*, vol. 61(10), pp. 2212–20.

Checkley, W., Buckley G., Gilman, R., Assis, A., Guerrant, R., Morris, S., Molbak, K., Valentiner-Branth, P., Lanata, C., Black, R. and The Childhood Malnutrition and Infection Network. (2008). 'Multi-country analysis of the effects of diarrhoea on childhood stunting', *International Journal of Epidemiology*, vol. 37(4), pp. 816–30.

Cohen, J. and Dupas, P. (2010). 'Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment', *Quarterly Journal of Economics*, vol. 125(1), pp. 1–45.

Conley, T. and Udry, C. (2010). 'Learning about a new technology: pineapple in Ghana', *American Economic Review*, vol. 100(1), pp. 35–69.

Curtis, V. and Cairncross, S. (2003). 'Effect of washing hands with soap on diarrhea risk in the community: a systematic review', *Lancet Infectious Diseases*, vol. 3(5), pp. 275–81.

De Walque, D. (2007). 'How does the impact of an HIV/AIDS information campaign vary with educational attainment? Evidence from rural Uganda', *Journal of Development Economics*, vol. 84(2), pp. 686–714.

Dupas, P. (2011). 'Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya', *American Economic Journal: Applied Economics*, vol. 3(1), pp. 1–34.

Ejemot-Nwadiaro, R., Ehiri, J., Meremikwu, M. and Critchley, J. (2008). 'Handwashing to prevent diarrhoea', *Cochrane Database of Systematic Reviews*, vol. 1.

Fenn, B., Bulti, A., Nduna, T., Duffield, A. and Watson, F. (2012). 'An evaluation of an operations research project to reduce childhood stunting in a food-insecure area in Ethiopia', *Public Health Nutrition*, vol. 15 (9), pp. 1746–54.

Fewtrell, L, Kaufmann, R.B., Kay, D., Enanoria, W., Haller, L. and Colford, J.M. (2005). 'Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis', *Lancet Infectious Diseases*, vol. 5(1), pp. 42–52.

Foster, A. and Rosenzweig, M. (1995). 'Learning by doing or learning from others: human capital and technical change in agriculture', *Journal of Political Economy*, vol. 103(6), pp. 1176–209.

Galiani, S. and Schargrodsky, E. (2004). 'Effects of land titling on child health', *Economics and Human Biology*, vol. 2(3), pp. 353–72.

Genius, M., Koundouri, P., Gauges, C. and Tzouvelekas, V. (2014). 'Information transmission in irrigation technology adoption and diffusion: social learning, extension services, and spatial effects', *American Journal of Agricultural Economics*, vol. 96(1), pp. 328–44.

Gertler, P. and Gruber, J. (2002). 'Insuring consumption against illness', *American Economic Review*, vol. 92(1), pp. 51–70.

Guerrant, D.I., Moore, S.R., Lima, A.A., Patrick, P.D., Schorling, J.B. and Guerrant, R.L. (1999). 'Association of early childhood diarrhea and cryptosporidiosis with impaired physical fitness and cognitive function four–seven years later in a poor urban community in Northeast Brazil', *American Journal of Tropical Medicine and Hygiene*, vol. 61(5), pp. 707–13.

Guiteras, R., Levine, D., Luby, S., Polley, T., Jannat, K.K. and Unicomb, L. (2014). 'Disgust, shame, and soapy water: tests of novel interventions to promote safe water and hygiene', Working Paper No. PWP-MPRC-2014-014, Maryland Population Research Center.

Halder, A., Tronchet, C., Akhter, S., Bhuiya, A., Johnston, R. and Luby, S. (2010). 'Observed hand cleanliness and other measures of hand-washing behavior in rural Bangladesh', *BMC Public Health*, vol. 10(1), pp. 545–64.

Heckman, J., Moon, S.H., Pinto, R., Savelyev, P. and Yavitz, A. (2010). 'Analyzing social experiments as implemented: a reexamination of the evidence from the HighScope Perry Preschool Program', *Quantitative Economics*, vol. 1(1), pp. 1–46.

Heckman, J., Pinto, R. and Savelyev, P. (2013). 'Understanding the mechanisms through which an influential early childhood program boosted adult outcomes', *American Economic Review*, vol. 103(6), pp. 1–35.

Hidrobo, M. (2014). 'The effect of Ecuador's 1999 economic crisis on early childhood development', *Economic Development and Cultural Change*, vol. 62(4), pp. 633–71.

Hunte, P. and Sultana, F. (1992). 'Health-seeking behavior and the meaning of medications in Balochistan, Pakistan', *Social Science and Medicine*, vol. 34(12), pp. 1385–97.

Jalan, J. and Somanathan, E. (2008). 'The importance of being informed: experimental evidence on the demand for environmental quality', *Journal of Development Economics*, vol. 87(1), pp. 14–28.

Kamyab, K., Nonnemaker, J. and Farrelly, M. (2014). 'Public support for graphic health warning labels in the U.S.', *American Journal of Preventative Medicine*, vol. 48(1), pp. 89–92.

Karmakar, P., Islam, M.M., Kibria, M.G., Hossain, M.S. and Sattar, M.M. (2012). 'Prevalence, belief and awareness of preferring traditional health care system in urban and rural people of Noakhali district, Bangladesh', *International Current Pharmaceutical Journal*, vol. 1(9), pp. 229–34.

Katz, L., Kling, J. and Liebman, J. (2001). 'Moving to opportunity in Boston: early results of a randomized mobility experiment', *Quarterly Journal of Economics*, vol. 116(2), pp. 607–54.

Kernan, W., Viscoli, C., Makuch, R., Brass, L. and Horwitz, R. (1999). 'Stratified randomization for clinical trials', *Journal of Clinical Epidemiology*, vol. 52(1), pp. 19–26.

Kooreman, P. and Baars, E. (2012). 'Patients whose GP knows complementary medicine tend to have lower costs and live longer', *European Journal of Health Economics*, vol. 13(6), pp. 769–76.

Kosek, M., Bern, C. and Guerrant, R. (2003). 'The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000', *Bulletin of the World Health Organization*, vol. 81(3), pp. 197–204.

Kremer, M., Leino, J., Miguel, E. and Zwane, A.P. (2011). 'Spring cleaning: rural water impacts, valuation, and property rights institutions', *Quarterly Journal of Economics*, vol. 126(1), pp. 145–205.

Kremer, M. and Miguel, E. (2007). 'The illusion of sustainability', *Quarterly Journal of Economics*, vol. 122(3), pp. 1007–65.

Krishnan, P. and Patnam, M. (2014). 'Neighbors and extension agents in Ethiopia: who matters more for technology adoption?', *American Journal of Agricultural Economics*, vol. 96(1), pp. 308–27.

Lamberti, L., Walker, C.F., Taneja, S., Mazumder, S. and Black, R. (2015). 'The influence of episode severity on caregiver recall, careseeking, and treatment of diarrhea among children 2–59 months of age in Bihar, Gujarat, and Uttar Pradesh, India', *American Journal of Tropical Medicine and Hygiene*, vol. 93(2), pp. 250–6.

Langford, R., Lunn, P. and Panter-Brick, C. (2011). 'Hand-washing, subclinical infections, and growth: a longitudinal evaluation of an intervention in Nepali Slums', *American Journal of Human Biology*, vol. 23(5), pp. 621–629.

Lee, S. and Shaikh, A. (2014). 'Multiple testing and heterogeneous treatment effects: re-evaluating the effect of PROGRESA on school enrollment', *Journal of Applied Econometrics*, vol. 29(4), pp. 612–26.

Leonard, K. (2003). 'African traditional healers and outcome-contingent contracts in health care', *Journal of Development Economics*, vol. 71(1), pp. 1–22.

Leonard, K. and Graff Ziven, J. (2005). 'Outcome versus service based payments in health care: lessons from African traditional healers', *Health Economics*, vol. 14(6), pp. 575–93.

Lin, M. and Liu, E. (2014). 'Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment', *Journal of Health Economics*, vol. 37, pp. 152–63.

Luby, S., Halder, A., Huda, T., Unicomb, L., and Johnston, R. (2011). 'Using child health outcomes to identify effective measures of hand-washing', *American Journal of Tropical Medicine and Hygiene*, vol. 85(5), pp. 882–92.

Luo, R., Shi, Y., Zhang, L., Zhang, H., Miller, G., Medina, A. and Rozelle, S. (2012). 'The limits of health and nutrition education: evidence from three randomized-controlled trials in rural China', *CESifo Economic Studies*, vol. 58(2), pp. 385–404.

Madajewicz, M., Pfaff, A., van Geen, A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R. and Ahsan, H. (2007). 'Can information alone change behavior? Response to arsenic contamination of ground-water in Bangladesh', *Journal of Development Economics*, vol. 84(2), pp. 731–54.

McKay, S., Gaudier, E., Campbell, D., Prentice, A. and Albers, R. (2010). 'Environmental enteropathy: new targets for nutritional interventions', *International Health*, vol. 2(3), pp. 172–80.

Mobarak, A.M., Dwivedi, P., Bailis, R., Hildemann, L. and Miller, G. (2012). 'Low demand for nontraditional cookstove technologies', *Proceedings of the National Academy of Sciences*, vol. 109(27), pp. 10815–20.

Mull, J.D. and Mull, D. (1988). 'Mothers' concepts of childhood diarrhea in rural Pakistan: what ORT program planners should know', *Social Science and Medicine*, vol. 27(1), pp. 53–67.

Niehaus, M., Moore, S., Patrick, P., Derr, L., Lorntz, B., Lima, A. and Guerrant, R. (2002). 'Early childhood diarrhea is associated with diminished cognitive function 4 to 7 years later in children in a northeast Brazilian Shantytown', *American Journal of Tropical Medicine and Hygiene*, vol. 66(5), pp. 590–3.

Nielsen, M., Hoogvorst, A., Konradsen, F., Mudasser, M. and der Hoek, W. (2003). 'Causes of childhood diarrhea as perceived by mothers in the Punjab, Pakistan', *Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 34(2), pp. 343–51.

Pattanayak, S., Yang, J., Dickinson, K., Poulos, C., Patil, S., Mallick, R., Blitstein, J. and Praharaj, P. (2009). 'Shame or subsidy revisited: social mobilization for sanitation in Orissa, India', *Bulletin of the World Health Organization*, vol. 87(8), pp. 580–87.

Ram, P. (2010). 'Practical guidance for measuring handwashing behavior', Working Paper, Water and Sanitation Program.

Ram, P., Jahid, I., Halder, A., Nygren, B., Islam, M.S., Granger, S., Molyneaux, J. and Luby, S. (2011). 'Variability in hand contamination based on serial measurements: implications for assessment of hand-cleansing behavior and disease risk', *American Journal of Tropical Medicine and Hygiene*, vol. 84(4), pp. 510–6.

Rhee, M., Mahamadou, S., Perry, S., McFalrnad, W., Parsonnet, J. and Ogobara, D. (2005). 'Use of insecticide-treated nets (ITNs) following a malaria education intervention in Piron, Mali: a control trial with systematic allocation of households', *Malaria Journal*, vol. 4(35), pp. 35–44.

Romano, J., Shaikh, A. and Wolf, M. (2008). 'Formalized data snooping based on generalized error rates', *Econometric Theory*, vol. 24(2), pp. 404–47.

Romano, J. and Wolf, M. (2005). 'Exact and approximate stepdown methods for multiple hypothesis testing', *Journal of the American Statistical Association*, vol. 100(469), pp. 94–108.

Ruel, M. and Arimond, M. (2002). 'Spot-check observational method for assessing hygiene practices: review of experience and implications for programmes', *Journal of Health, Population, and Nutrition*, vol. 20(1), pp. 65–76.

Sato, A. (2012a). 'Do inequalities in health care utilization in developing countries change when we take into account traditional medicines?', *World Development*, vol. 40(11), pp. 2275–89.

Sato, A. (2012b). 'Does socio-economic status explain the use of modern and traditional care services?', *Social Science and Medicine*, vol. 75(8), pp. 1450–9.

Schmidt, W., Arnold, B., Boisson, S., Genser, B., Luby, S., Barreto, M., Clasen, T. and Cainrcross, S. (2011). 'Epidemiological methods in diarrhoea studies – an update', *International Journal of Epidemiology*, vol. 40 (6), pp. 1678–92.

Schmidt, W., Genser, B., Barreto, M., Clasen, T., Luby, S., Cainrcross, S. and Chalbi, Z. (2010). 'Sampling strategies to measure the prevalence of common recurrent infections in longitudinal studies', *Emerging Themes in Epidemiology*, vol. 7(15), pp. 1–13.

Sheehan, H. and Hussain, S.J. (2002). 'Unani tibb: history, theory, and contemporary practice in South Asia', *Annals of the American Academy of Political and Social Science*, vol. 583(1), pp. 122–35.

Tenkorang, E., Gyimah, S., Maticka-Tyndale, E. and Adjei, J. (2011). 'Superstition, witchcraft, and HIV prevention in sub-Saharan Africa: the case of Ghana', *Culture, Health, and Sexuality*, vol. 13(9), pp. 1001–14.

Thorton, R. (2008). 'The demand for, and impact of, learning HIV status', *American Economic Review*, vol. 98 (5), pp. 1829–63.

Vindigni, S., Riley, P. and Jhung, M. (2011). 'Systematic review: hand-washing behaviour in low- to middle-income countries: outcome measures and behaviour maintenance', *Tropical Medicine and International Health*, vol. 16(4), pp. 466–77.

Waddington, H. and Snilstveit, B. (2009). 'Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea', *Journal of Development Effectiveness*, vol. 1(3), pp. 295–335.

Wang, W., Keh, H.T., and Bolton, L. (2010). 'Lay theories of medicine and a healthy lifestyle', *Journal of Consumer Research*, vol. 37, pp. 80–97.

Wishnofsky, M. (1958). 'Caloric equivalents of gained or lost weight', *American Journal of Clinical Nutrition*, vol. 6(5), pp. 542–6.

Wong, H.L., Shi, Y., Luo, R., Zhang, L. and Rozelle, S. (2014). 'Improving the health and education of elementary schoolchildren in rural China: iron supplementation versus nutritional training for parents', *Journal of Development Studies*, vol. 50(4), pp. 502–19.

World Health Organization. (2003). 'Traditional medicine', Fact Sheet No. 134.

World Health Organization. (2013a). 'Diarrhoeal disease', Fact Sheet No. 330.

World Health Organization. (2013b). 'WHO traditional medicine strategy: 2014–2023', Technical Report.

World Health Organization, Global Health Workforce Alliance. (2008). 'Country case study: Pakistan's lady health worker programme', Technical Report.

Yamba, C.B. (1997). 'Cosmologies in turmoil: witchfinding and AIDS in Chiawa, Zambia', *Africa: Journal of the International African Institute*, vol. 67(2), pp. 200–23.

Zafar, S.N., Luby, S.P. and Mendoza, C. (2010). 'Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall', *Epidemiology and Infection*, vol. 138(2), pp. 264–9.